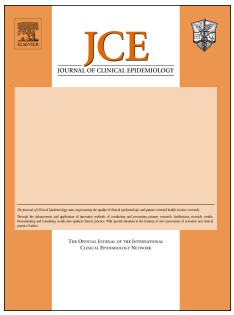


Methods used in the selection of instruments for outcomes included in Core Outcome Sets have improved since the publication of the COSMIN/COMET guideline

Sarah L. Gorst, Cecilia AC. Prinsen, Maximilian Salcher-Konrad, Karen Matvienko-Sikar, Paula R. Williamson, Caroline B. Terwee

PII: S0895-4356(20)30106-2

DOI: https://doi.org/10.1016/j.jclinepi.2020.05.021

Reference: JCE 10165

To appear in: *Journal of Clinical Epidemiology*

Received Date: 7 February 2020

Revised Date: 21 April 2020

Accepted Date: 20 May 2020

Please cite this article as: Gorst SL, Prinsen CA, Salcher-Konrad M, Matvienko-Sikar K, Williamson PR, Terwee CB, Methods used in the selection of instruments for outcomes included in Core Outcome Sets have improved since the publication of the COSMIN/COMET guideline, *Journal of Clinical Epidemiology* (2020), doi: https://doi.org/10.1016/j.jclinepi.2020.05.021.

Methods used in the selection of instruments for outcomes included in Core Outcome Sets have improved since the publication of the COSMIN/COMET guideline

Sarah L Gorst[a]; Cecilia AC Prinsen[b]; Maximilian Salcher-Konrad[c]; Karen Matvienko-Sikar[d]; Paula R Williamson[a], Caroline B Terwee[b]

Affiliations:

[a] Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom.

[b] Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands.

[c] Department of Health Policy, London School of Economics and Political Science, Cowdray House, London, United Kingdom.

[d] School of Public Health, University College Cork, Cork, Ireland.

Funding: This work was supported by the European Union's Horizon 2020 research and innovation program (CORBEL under grant agreement number 654248).

Conflict of interest statement: PRW is a member of the COMET (Core Outcome Measures in Effectiveness Trials) Management Group. CBT is founder of the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) initiative. The remaining authors declare no conflict of interest.

Corresponding author:

Dr Sarah Gorst

Department of Biostatistics,

University of Liverpool,

Block F Waterhouse Building,

1-5 Brownlow Street,

Liverpool, L69 3GL,

United Kingdom.

Email: sgorst@liverpool.ac.uk

Tel: +44 (0) 151 794 9725

Abstract

Objectives: Once a core outcome set (COS) has been defined, it is important to achieve consensus on how these outcomes should be measured. The aim of this systematic review is to gain insight into the methods used to select outcome measurement instruments and determine whether methods have improved following the COSMIN/COMET guideline publication.

Study Design and Setting: Eligible articles, which were identified from the annual COMET systematic review, concerned any COS development studies that provided a recommendation on how to measure the outcomes included in the COS. Data was extracted on the methods used to select outcome measurement instruments in accordance with the COSMIN/COMET guideline.

Results: Of the 118 studies included in the review, 48% used more than one source of information when finding outcome measurement instruments and 74% performed some form of quality assessment of the measurement instruments. Twenty-three studies recommended one single instrument for each core outcome included in the COS. Clinical experts and public representatives were involved in selecting instruments in 62% and 28% of studies, respectively.

Conclusion: Methods used to select outcome measurement instruments have improved since the publication of the COSMIN/COMET guideline. Going forward COS developers should ensure that recommended outcome measurement instruments have sufficient content validity. In addition, COS developers should recommend one instrument for each core outcome to contribute to the overarching goal of uniformity in outcome reporting.

Key words: Core Outcome Set; COS; development; outcome measurement instrument; selection; methodology.

Running title: Methods used in selection of outcome measurement instruments.

Word count: 3929 words

What is new?

Key findings

- Methods used to select core outcome measurement instruments vary across studies, with many studies not meeting the recommended standards.
- Methods used to select outcome measurement instruments have improved since the publication of the COSMIN/COMET guideline.

What this adds to what is known?

- This is the first study to assess how the outcome measurement instruments recommended in existing core outcome sets have been selected and whether good practices are being followed.

What is the implication, what should change now?

- Core outcome set developers need to make better use of the guidance available when agreeing on how to measure the outcomes included in core outcome sets.
- Developers need to ensure that outcome measurement instruments are of sufficient quality and especially have sufficient content validity.

1. Introduction

There is lack of consensus with regard to the selection of outcomes and outcome measurement instruments for clinical trials, which causes inconsistencies in the outcomes reported and difficulties in comparing these outcomes in systematic reviews and meta-analyses.[1] In addition, there is great variability in the quality of outcome measurement instruments used, and it is not always clear if the best instrument is being used for a given outcome. To overcome these issues, standardisation of the selection of outcomes and outcome measurement instruments is needed.

The Core Outcome Measures in Effectiveness Trials (COMET) Initiative (www.comet-initiative.org), launched in January 2010, aims to facilitate the development and application of agreed standardized sets of outcomes, also known as 'Core Outcome Sets' (COS). A COS is an agreed minimum set of outcomes that should be measured and reported in all clinical trials of a specific disease or trial population (i.e. *what* to measure).[1] Once the COS has been defined, it is then important to achieve consensus on how these outcomes should be measured (i.e. *how* to measure).

The COSMIN initiative (COnsensus-based Standards for the selection of health Measurement Instruments, http://www.cosmin.nl/) aims to improve the selection of outcome measurement instruments.[^] In 2016, COSMIN and COMET published a consensus-based guideline on how to select outcome measurement instruments for outcomes included in a COS.[2] Among a large group of international stakeholders from 14 different countries, including clinicians, clinimetricians/psychometricians, epidemiologists, journal editors, physicians, researchers, and statisticians, consensus was obtained on methods for selecting outcome measurement

[^]When using the term 'outcome measurement instruments' we are referring to any instruments, definitions, tools, procedures, etc., that are used to measure an outcome.

instruments for outcomes included in a COS. COS developers are guided through the process of instrument selection in four consecutive steps: Step 1, conceptual considerations; Step 2, finding existing outcome measurement instruments; Step 3, quality assessment of outcome measurement instruments; Step 4, recommendations on the selection of outcome measurement instruments. It is unknown, however, how the outcome measurement instruments recommended in existing COS have actually been selected by COS developers and whether good practices for COS development, as described in the COSMIN/COMET guideline are being followed.

The aim of this systematic review is to: (1) gain insight into the methods used by COS developers for selecting core outcome measurement instruments; and (2) determine whether the methods have improved following the 2016 publication of the COSMIN/COMET guideline. It is hypothesised that the quality of the methods used to select the core outcome measurement instruments varies considerably, and it is therefore anticipated that there will be considerable room for improvement in COS development with regard to instrument selection. However, there are expected to be improvements in the methods used in studies that have been published from 2017 onwards.

2. Methods

2.1. Eligibility criteria

Eligible articles concern COS development studies that provide recommendations on what and how to measure, either done together in one study or done in two separate stages (i.e. two or more studies). We included all COS studies, identified from the original COMET systematic review and annual updates,[3-8] that provided a recommendation on the

instruments to measure the outcomes included in the COS. COS development studies that only provide a recommendation on what to measure, but do not consider how to measure, and studies that discuss how to measure the outcomes but do not give a recommendation were excluded.

2.2. Literature search

The search strategy for identifying eligible COS development studies has been described elsewhere in detail.[3-8] In brief, a comprehensive search strategy to identify studies that aimed to define COS in any disease area was first developed in 2013[3] (see Appendix A for full search strategy). Database searches were repeated in 2015,[4] 2016,[5], 2017[6], 2018[7] and 2019.[8]

2.3. Data extraction

Four sets of reviewers (CP-SG, SG-MSK, MSK-CP, SG-KMS) independently extracted data on descriptive information for each identified COS, including the target population, disease area, and the (number of) outcomes. Data has been extracted in accordance with the COSMIN/COMET guideline (see flowchart in Appendix B). Methods used to select instruments for the COS have been extracted, including the approach taken to identify existing instruments, the evaluation of the quality and feasibility of instruments, the number of instruments recommended for use, arguments used for selecting instruments other than quality criteria (measurement properties), recommendations for additional research on instruments, and whether any guidance for instrument selection, including the COSMIN/COMET guideline, was followed (see Appendix C). To ensure consistency in data extraction, the data extraction form was pilot tested for a set of five studies and the extracted data was compared before extracting data for the remaining studies. Discrepancies in data

extraction between pairs of reviewers were sought to be resolved by discussion with the third reviewer and consensus was reached.

To improve the quality of our data, first authors of the included studies were contacted in person by email to verify the data extracted from their studies and they were asked to provide additional information that might be missing. In case the email could not be delivered, the last author of that particular study was contacted by email. References of the included papers were also checked to identify any other relevant articles on instrument selection for COS.

3. Results

A total of 163 articles describing 118 COS development studies were included in the review. A flow diagram of the article and abstract selection process is provided in Figure 1, guided by PRISMA.[9]

Details on COS development studies (e.g. target population, disease area) can be found in Appendix D. In summary, COS were developed in a variety of geographical locations, including Asia, Canada, Europe, New Zealand, South Africa, South America, and USA. All COS were developed in the English language. The number of core outcomes included in the COS varied between two[10, 11] and 26.[12] Following the COMET classifications[13] COS were developed in 24 different disease areas, mostly neurology (n=19 studies), rheumatology (n=14), heart & circulation (n=13), and orthopaedics & trauma (n=11). In 36 studies, the COS were developed for adults; in 17 studies for children and in 20 studies for both adults and children. In 45 studies (38%), the age group for which the COS was developed was not specified. Of the 118 studies, 23 studies reported on core outcomes for different subgroups of patients, such as age groups;[14-16] acute or chronic conditions;[17] disease severity;[18, 19]

type of study, for example prevention trials vs intervention trials[20, 21] or phase I-II vs phase III clinical trials;[15] for acute vs long-term treatment[22] or acute treatment vs prophylaxis;[23-25] for different diseases/conditions;[16, 26-32] or for different settings.[33, 34]

3.1. Methods used in the selection of instruments for COS

With regard to COS development, 87/118 COS studies used a single process to identify the core outcomes as well the instruments recommended to measure these outcomes; whereas 31 studies used a two stage process that involved first agreeing on 'what to measure' (select core outcomes) before moving onto the 'how to measure' (recommend instruments).

3.1.1. Finding existing outcome measurement instruments

It is recommended that COS developers aim for finding all existing outcome measurement instruments. Multiple sources of information can be used to find instruments: (1) performing a systematic review, including a search in MEDLINE (and EMBASE); (2) use existing review(s); (3) reference lists; (4) expert opinion; or (5) other sources of information, such as online databases, book (chapters) or conference proceedings.[2] Of the 118 included studies, 21 studies (18%) used three or more sources of information when finding existing outcome measurement instruments and 36 studies (30%) used two sources of information. Sixty-one studies (52%) used only one source of information when finding existing outcome measurement instruments, with 39/61 studies accessing expert opinion only. Of the 118 studies, 52 studies (44%) performed a systematic review; 19 studies (16%) used an existing review; seven studies (6%) searched reference lists; 49 studies (42%) accessed expert opinion; and 15 studies (13%) used other sources of information, mostly instruments used in clinical trials.

3.1.2. Quality assessment of outcome measurement instruments

COS developers should base their recommendations for outcome measurement instruments on (1) the quality of the existing outcome measurement instruments, i.e. their measurement properties (including an evaluation of the quality of the validation studies), and (2) the feasibility aspects of the outcome measurement instruments.[2]

3.1.2.1. Measurement properties

It is recommended that evidence on the measurement properties of outcome measurement instruments should be available in the target population. The quality of instruments is determined in studies on measurement properties, which should be of high methodological quality. Of the 118 studies, six studies[35-40] (5%) considered both the results of the measurement properties of the outcome measurement instruments and the quality of studies on these measurement properties. However, in three of these studies[35, 36, 38] it remains unclear whether a best evidence synthesis was performed. COS developers of 23/118 studies (20%) only considered the results of the measurement properties of the included outcome measurement instruments, but did not consider the quality of the studies on these measurement properties. In 58/118 studies (49%), COS developers referred to quality criteria of outcome measurement instruments, however there was no mention of any formal assessment of whether the instruments met these criteria. COS developers of 31/118 studies (26%) did not take the quality of the outcome measurement instruments into account when making their recommendations.

3.1.2.2. Feasibility

Of the 118 studies, 74 studies (63%) have taken feasibility aspects into consideration in the selection of instruments for the COS, such as availability of the instrument, cost of an instrument, ease of administration, and length of an instrument. In 44 studies (37%) there was no indication that feasibility was taken into consideration in the selection of instruments for the COS.

3.1.3. Recommendations on the selection of outcome measurement instruments

It is advised to recommend only one outcome measurement instrument for each core outcome per subdomain/subpopulation in the COS, as this will serve the ultimate goal of standardization of outcome reporting.[2] Of the 118 studies, only 11 studies (9%) recommended one single instrument for each core outcome included in the COS.[12, 38, 39, 41-48] In seven of these 11 studies, one instrument was selected for each core outcome (range: 4-26) in the COS.[12, 38, 41, 43, 44, 47, 48] In three studies, one instrument was recommended for each core outcome or for each subpopulation (i.e. children and adolescents) (range of core outcomes: 4-10).[42, 45, 46] In one study, one instrument was recommended for each of the four core outcomes, with two alternative instruments recommended for two of the outcomes because they were free of charge.[39] In 12 of the 118 studies (10%), one single instrument was recommended for each core outcome other than those for which no outcome measurement instrument could be recommended (range of core outcomes with instrument recommended: 1-17).[21, 37, 49-58] Twelve of the 118 studies (10%) recommended one instrument for all included core outcomes except one (range included core outcomes: 3-15).[25, 35, 40, 59-67] Another seven studies (6%) recommended multiple instruments for all core outcomes included in the COS.[19, 36, 68-72] In 76/118 studies (64%) a combination of recommendations was used in the selection of instruments for each core outcome included in the COS (i.e. for some outcomes one instrument was

recommended, for some outcomes multiple instruments were recommended, either for the entire group or for different subgroups and for some outcomes no instrument was recommended).

3.1.4. Consensus procedure used to reach agreement

It is recommended that COS developers use a consensus procedure to get final agreement on the selected instruments included in the COS.[2] In 80/118 studies (68%) a consensus procedure was used. In 30/80 studies it was unclear and not specified how consensus was obtained. In 8/80 studies COS developers used a Delphi technique to reach consensus on the selection of core instruments.[12, 43, 48, 64, 73-76] Six of the 80 studies were guided by the OMERACT consensus and validation process, which involved participants voting and then breaking out into groups to review and discuss domains and instruments.[39, 40, 63, 72, 77, 78] In 36/80 studies COS developers conducted a consensus meeting, including various methods, to reach consensus on the core instruments. Consensus methods used at the meetings included presentations, nominal group techniques, group discussions, consensus workshops, breakout sessions and voting. In 38/118 studies (32%), no consensus procedure was used to agree on the instruments included in the COS and recommendations were formulated by the COS developers.

3.1.4. Stakeholders involved in the selection of outcome measurement instruments

Of the 118 studies, the following stakeholders were involved in the selection of outcome measurement instruments: clinical experts (n=73), non-clinical researchers (n=39), patients and/or public representatives (n=33), regulatory authorities (n=20) and industry representatives (n=17). An additional 43 studies did not provide any details about the

stakeholders involved in selecting the outcome measurement instruments. Table 1 displays the different stakeholder combinations across the 118 studies.

Table 1. Number of studies involving each stakeholder group combination

Stakeholder groups	n (%)
Clinical experts	19 (16)
Clinical experts, public representatives and non-clinical research experts	11 (9)
Clinical experts and non-clinical research experts	8 (7)
Clinical experts and public representatives	7 (6)
Clinical experts, public representatives, non-clinical research experts and industry experts	6 (5)
Clinical experts, public representatives, non-clinical research experts, authorities and industry experts	4 (3)
Clinical experts, public representatives, non-clinical research experts and authorities	3 (3)
Clinical experts, non-clinical research experts and authorities	3 (3)
Clinical experts, authorities and industry experts	3 (3)
Clinical experts, public representatives and authorities	2 (2)
Clinical experts and authorities	2 (2)
Clinical experts and other	2 (2)
Clinical experts, non-clinical research experts, authorities and industry experts	2 (2)
Clinical experts and industry experts	1 (1)
Non-clinical research experts, authorities and industry experts	1 (1)
Non-clinical research experts	1 (1)
No details provided	43 (36)

3.1.5. Guidance on instrument selection

In 35/118 studies (30%) published guidance for instrument selection was used. Most studies (n=13) used the OMERACT guidance,[79] whereas other studies used the Grading of Recommendations, Assessment, and Evaluation (GRADE) approach;[16, 80] IMMPACT recommendations;[81, 82] WHO-ICF framework;[83, 84] EULAR operating procedures;[85] ICHOM framework.[12, 48, 64] Four studies used the COSMIN/COMET guideline[39, 40, 76, 84] and a fifth study[37] used the COSMIN standards for the selection of health status measurement instruments.[86] Eleven studies[11, 17, 32, 36, 87-93] referred to other guidance, for example guidelines from ECCO GuiCom; guidance by Physical Rehabilitation

Outcomes Measures, published by the Canadian Physiotherapy Association; previous work by Hudak et al[94] or Dworkin et al;[95] other consensus guidelines;[96-101] the HOME roadmap;[102] the framework for the selection of clinical trial indices proposed by Tugwell and Bombardier;[103] and guidance from a qualitative evaluation of measures for psychosocial intervention in dementia care. No guidance was used in 43/118 studies (37%), whereas in 40/118 studies (34%) it remains unclear whether any form of guidance was used.

3.1.6. Recommendations for additional research on instruments

In 55/118 studies (47%) recommendations were made for additional validation studies (n=31), or development of new instruments (n=24).

3.2. Differences between studies published before and after the publication of the COSMIN/ COMET guideline

Of the 118 studies included in this review, 92 studies (78%) were published prior to the publication of the COSMIN/COMET guideline and 26 studies (22%) were published following its publication. Table 2 provides a comparison of the methods used in studies published before and after the publication of the COSMIN/COMET guideline. Studies published following the publication of the COSMIN/COMET guideline were more likely to base their recommendations on the quality of the outcome measurement instruments, with 8/26 studies (31%) considering the evaluation of the measurement properties of the instruments and a further 4/26 studies (15%) considering both the quality of the measurement properties of the outcome measurement instruments and the quality of studies on measurement properties. Additionally, these studies were also more likely to comply with the COSMIN/COMET recommendations on the selection of outcome measurement instruments, with 11/26 studies (42%) recommending one single instrument for each core outcome

included in the COS, for which a recommendation could be made, and a further 6/26 studies (23%) recommending one instrument for all included core outcomes except one. The inclusion of stakeholders across all groups increased in the post-guideline studies, with the biggest increase being the inclusion of patients and/or public representatives, which increased from 20% to 58%. There was also a 31% decrease in the number of studies that did not provide any details about stakeholder involvement. Regarding the use of guidance, 12/26 studies (46%) published following the COSMIN/COMET guideline used available guidance for instrument selection, with 5/26 studies (19%) specifically using the COSMIN/COMET guideline or other COSMIN guidance.

Table 2. Methods used in studies published before and after the publication of the COSMIN/COMET guideline

	Pre-guideline n (%)	**Post-guideline n (%)**
Finding existing outcome measurement instruments		
Used 3 or more sources to find existing outcome measurement instruments	16/92 (17)	5/26 (19)
Used 2 sources to find existing outcome measurement instruments	30/92 (33)	6/26 (23)
Used 1 source to find existing outcome measurement instruments	46/92 (50)	15/26 (58)
Quality assessment of outcome measurement instruments		
(a) Measurement properties		
Considered both the quality of the measurement properties of the outcome measurement instruments and the quality of studies on measurement properties	2/92 (2)	4/26 (15)
Considered the evaluation of the measurement properties of the included outcome measurement instruments	15/92 (16)	8/26 (31)
Referred to quality criteria but no formal assessment	52/92 (57)	6/26 (23)
Quality of outcome measurement instruments not taken into account	23/92 (25)	8/26 (31)
(b) Feasibility		
Feasibility aspects taken into consideration	59/92 (64)	15/26 (58)
Recommendations on the selection of outcome measurement instruments		
One single instrument for each core outcome	3/92 (2)	8/26 (31)
One single instrument for each core outcome where a recommendation could be made	9/92 (10)	3/26 (12)
One instrument for all included core outcomes except one	6/92 (7)	6/26 (23)
Multiple instruments for all core outcomes	7/92 (8)	0/26 (0)
Combination of recommendations for each core outcome	67/92 (73)	9/26 (35)
Consensus procedure used to reach agreement		
Used consensus procedure to get final agreement on the selected instruments	62/92 (67)	18/26 (69)

Specified details of the consensus procedure	35/62 (56)	15/18 (83)
Stakeholders involved in the selection of outcome measurement instruments		
Clinical experts	50/92 (54)	23/26(88)
Patients and/or public representatives	18/92 (20)	15/26 (58)
Non-clinical researchers	26/92 (28)	13/26 (50)
Regulatory authorities	15/92 (16)	5/26 (19)
Industry representatives	13/92 (14)	4/26 (15)
No details provided about the stakeholders involved in selecting the outcome measurement instruments	40/92 (43)	3/26 (12)
Guidance on instrument selection		
Used published guidance	23/92 (25)	12/26 (46)
Recommendations for additional research on instruments		
Made recommendations for the development of new instruments or additional validation studies	42/92 (46)	13/26 (50)

4. Discussion

We identified 118 COS development studies that provided recommendations for how to measure the outcomes included in a COS. Reviewing these studies has enabled us to gain insight into the methods used by COS developers to select outcome measurement instruments. Following the publication of the COSMIN/COMET guideline, there has been an improvement in the methods used, specifically in relation to quality assessment, recommendations on the selection of instruments, stakeholder involvement, and the use of published guidance.

4.1. Finding all existing instruments

COS developers should make better use of the literature to inform their instrument selection process. Relying solely on expert opinion to find existing outcome measurement instruments, as 33% of studies did in the current review, may result in only the most commonly used instruments, or those that are favored by clinicians, being considered. To assist developers in identifying instruments, COSMIN maintains a database of systematic reviews of outcome measurement instruments.[104] The 'COSMIN guideline for systematic reviews of patient-reported outcome measures'[105] can be used for performing a comprehensive literature

search or full systematic review to find all available instruments if a good quality systematic review is not available.

4.2. Quality assessment of instruments

COS developers should take both the quality of the studies on measurement properties and the results of the measurement properties of the outcome measurement instruments into account in their recommendations. This will ensure that the most reliable and valid outcome measurement instruments are selected. However, only ten of the COS development studies included in this review considered both aspects. To assist in the assessment of the quality of the studies, COSMIN has developed a Risk of Bias checklist for use in systematic reviews of PROMs to assess the risk of bias of studies on measurement properties.[106] COSMIN has also proposed quality criteria for measurement properties of health status questionnaires to assist COS developers in assessing the quality of identified instruments.[106]

In contrast to the assessment of measurement properties, the feasibility of the identified instruments was generally taken into consideration in the selection of outcome measurement instruments. We expect that the number of studies considering these feasibility aspects will continue to remain high, following the publication of the COSMIN/COMET guideline in 2016.[2]

4.3. Generic recommendations on the selection of outcome measurement instruments

COS developers should try to recommend only a single instrument for each individual outcome included in a COS, to contribute to the overarching goal of uniformity in outcome reporting and enhance the comparability of clinical trials. Exceptions can however be made

for subpopulations, such as children and adults, were different instruments may be necessary for the different age groups.

COS stakeholder groups should include more representative stakeholders, including patients, when agreeing on the most appropriate outcome measurement instruments. Most COS stakeholder groups were comprised of clinical experts, while public representatives were involved in less than one third of studies. This is concerning, as it suggests that the outcome measures recommended may not be those that public representatives deem to be most appropriate. Patients are increasingly being included in selecting the outcomes for inclusion in COS, with 92% of ongoing COS development studies in the COMET database planning to include some degree of patient input.[107] However, patients and/or public representatives were only included in the selection of outcome measurement instruments in 28% of COS studies.

It could be argued that it is more difficult to include patients in the selection of outcome measurement instruments than in the selection of the outcomes, because the selection of instruments is mostly based on studies on measurement properties, which may be difficult for patients to understand. COS developers may need to address additional issues when the population concerned includes people with cognitive impairment, communication difficulties or other vulnerabilities, which make participation in such processes challenging. Despite the complexities involved in the selection of outcome measurement instruments, there has been an increase in the inclusion of patients and/or public representatives since the publication of the COSMIN/COMET guideline. Thus, it is certainly possible to involve patients and/or public representative, including those from vulnerable groups, in this process. However, it should be acknowledged that these groups are likely to need additional support to participate

and so in some instances a multi-stage approach, which allows for adequate support, might be necessary to ensure that engagement is meaningful. One potential means of involvement is for patients to judge the face and content validity (relevance, comprehensiveness and comprehensibility) of the available outcome measurement instruments, which are considered the most important measurement properties.[106, 108] However, it should be noted that this may not be necessary, if previous published studies have assessed the validity of outcome measurement instruments with the population in question.

4.4. Implications

This review has highlighted that the methods used to select the core outcome measurement instruments vary across studies, with many studies not meeting the recommended standards. However, the majority of included studies were published prior to the development of the COSMIN/COMET guideline, in 2016, and so developers may have been unaware of methodology for selecting outcome measurement instruments. There have however been clear improvements in the methods used to select outcome measurement instruments in studies published since the publication of the COSMIN/COMET guideline. It is unclear whether such improvements are a direct result of the publication of the COSMIN/COMET guideline or whether other variables are responsible for the pre- and post-guideline reporting differences. Other potential variables may include increased COS awareness prompting COS developers to be more thorough in the outcome measurement instrument selection process. In addition, the differences in the rate of inclusion of patients may be attributable to increased public input in health research in general.

Going forward we hope that COS developers will utilise the COSMIN/COMET guideline, along with the other resources listed above, to ensure that recommendations for outcome

measurement instruments are developed using rigorous methodology. A recent paper by Ju and colleagues[109] highlights how to apply the COSMIN/COMET guidance when identifying outcome measurement instruments.

Apart from the COSMIN/COMET guideline, other guidelines can also be used to guide the selection of outcome measurement instruments for COS, for example the OMERACT Handbook[110] or HOME roadmap.[102] Different guidelines put different emphasis on different steps of the process. For example, in contrast to the COSMIN/COMET guideline and HOME roadmap, the OMERACT process does not require a search to find ALL available instruments but starts with a selection of instruments that seem to have good match with the target domain and are considered feasible. Both the COSMIN/COMET guideline and the OMERACT Handbook address the need for good content validity of outcome measurement instruments. Therefore, we suggest that, when resources are limited, COS developers should evaluate the content validity of available instruments, if this has not been done previously, with a small number of patients in their stakeholder group, e.g. ask patients to evaluate the relevance, comprehensiveness and comprehensibility of all items/tests. The COSMIN methodology for assessing content validity of PROMs can be used for further guidance.[108] Nevertheless, it should be acknowledged that while guidelines are important, some flexibility should be employed to best facilitate the participation of patients and/or public representatives.

4.5. Limitations

All of the studies included in the current review were identified from the annual COMET systematic review of COS. We did not perform a systematic search for all studies relating to how outcomes should be measured. For example, COS groups may perform systematic

reviews of outcome measurement instruments and select their instruments based on these reviews, but may not publish a separate paper on the selection process for the outcome measurement instruments; therefore, we cannot be certain that we have identified all relevant studies. However, we did check the references of all included papers and authors of COS studies were contacted to provide additional information.

4.6. Conclusions

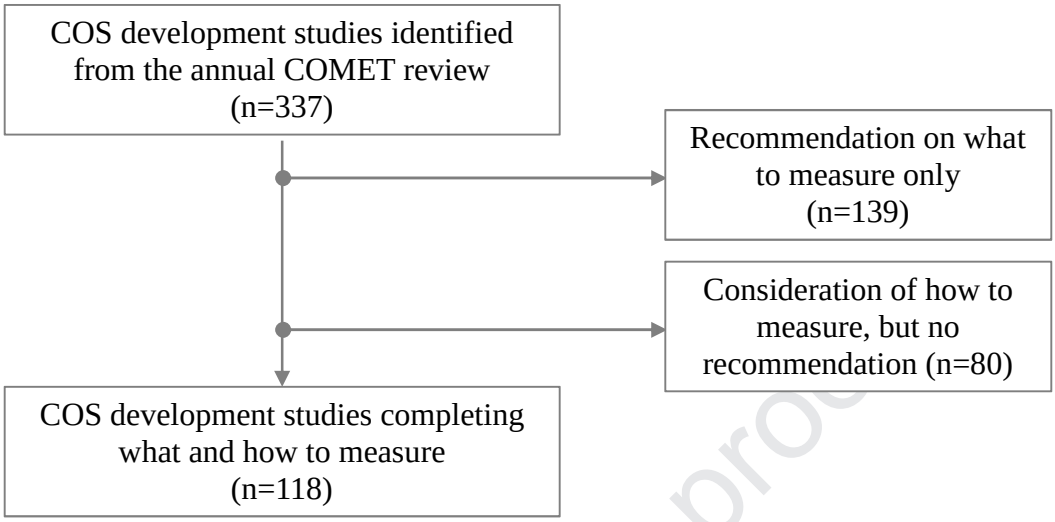
In conclusion, COS developers need to make better use of the guidance available when agreeing on how to measure the outcomes included in COS. Specifically, developers need to ensure that outcome measurement instruments are of sufficient quality and especially have sufficient content validity. Furthermore, developers should aim to adhere to uniformity by selecting a single outcome measurement instrument for each outcome within a COS.

Acknowledgements

The authors would like to thank the reviewers for their helpful comments on the manuscript.

418 **Figure 1. PRISMA flow chart of identification of eligible studies from the COMET**

419 **database**



420

References

1. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13:132.
2. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials. 2016;17(1):449.
3. Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. PLoS One. 2014;9(6):e99111.
4. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. PLoS One. 2016;11(1):e0146444.
5. Gorst SL, Gargon E, Clarke M, Smith V, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and Identification of Gaps. PLoS One. 2016;11(12):e0168403.
6. Davis K, Gorst SL, Harman N, Smith V, Gargon E, Altman DG, et al. Choosing important health outcomes for comparative effectiveness research: An updated systematic review and involvement of low and middle income countries. PLoS One. 2018;13(2):e0190695.
7. Gargon E, Gorst SL, Harman NL, Smith V, Matvienko-Sikar K, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 4th annual update to a systematic review of core outcome sets for research. PLoS One. 2018;13(12):e0209869.
8. Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. PLoS One. 2019;14(12):e0225980.
9. PRISMA Statement. http://www prisma-statement org/ [Internet]. 2016 10/27/2016. Available from: http://www.prisma-statement.org/.
10. Donovan DM, Bigelow GE, Brigham GS, Carroll KM, Cohen AJ, Gardin JG, et al. Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection and measurement of drug use end-points in clinical trials. Addiction. 2012;107(4):694-708.
11. Keim SM, Spaite DW, Maio RF, Garrison HG, Desmond JS, Gregor MA, et al. Risk adjustment and outcome measures for out-of-hospital respiratory distress. Acad Emerg Med. 2004;11(10):1074-81.
12. Ong WL, Schouwenburg MG, van Bommel ACM, Stowell C, Allison KH, Benn KE, et al. A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative. JAMA Oncol. 2017;3(5):677-85.
13. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol. 2017.
14. Busse WW, Morgan WJ, Taggart V, Togias A. Asthma outcomes workshop: overview. J Allergy Clin Immunol. 2012;129(3 Suppl):S1-8.
15. Ramsey BW, Boat TF. Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference. Journal of Pediatrics. 1994;124(2):177-92.
16. Sanyal AJ, Brunt EM, Kleiner DE, Kowdley KV, Chalasani N, Lavine JE, et al. Endpoints and clinical trial design for nonalcoholic steatohepatitis. Hepatology. 2011;54(1):344-53.
17. McGrath PJ, Walco GA, Turk DC, Dworkin RH, Brown MT, Davidson K, et al. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. J Pain. 2008;9(9):771-83.
18. Vellas B, Andrieu S, Sampaio C, Coley N, Wilcock G. Endpoints for trials in Alzheimer's disease: a European task force consensus. The Lancet Neurology. 2008;7(5):436-50.
19. Chen RC, Chang P, Vetter RJ, Lukka H, Stokes WA, Sanda MG, et al. Recommended patient-reported core set of symptoms to measure in prostate cancer treatment trials. J Natl Cancer Inst. 2014;106(7).

20. Abellan van Kan G, Cameron Chumlea W, Gillette-Guyonet S, Houles M, Dupuy C, Rolland Y, et al. Clinical trials on sarcopenia: methodological issues regarding phase 3 trials. Clinics in Geriatric Medicine. 2011;27(3):471-82.
21. Balakrishnan K, Sidell DR, Bauman NM, Bellia-Munzon GF, Boesch RP, Bromwich M, et al. Outcome measures for pediatric laryngotracheal reconstruction: International consensus statement. Laryngoscope. 2019;129(1):244-55.
22. Recommendations for guidelines on clinical trials of mucoactive drugs in chronic bronchitis and chronic obstructive pulmonary disease. Task Group on Mucoactive Drugs. Chest. 1994;106(5):1532-7.
23. Tfelt-Hansen P, Pascual J, Ramadan N, Dahlof C, D'Amico D, Diener HC, et al. Guidelines for controlled trials of drugs in migraine: third edition. A guide for investigators. Cephalalgia. 2012;32(1):6-38.
24. Lipton RB. Guidelines for controlled trials of drugs in cluster headache. Cephalalgia. 1995;15(6):452-62.
25. Bendtsen L, Bigal ME, Cerbo R, Diener HC, Holroyd K, Lampl C, et al. Guidelines for controlled trials of drugs in tension-type headache: second edition. Cephalalgia. 2010;30(1):1-16.
26. Ruperto N, Ravelli A, Murray KJ, Lovell DJ, Andersson-Gare B, Feldman BM, et al. Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis. Rheumatology (Oxford). 2003;42(12):1452-9.
27. Porst H, Vardi Y, Akkus E, Melman A, Park NC, Seftel AD, et al. Standards for clinical trials in male sexual dysfunctions. J Sex Med. 2010;7(1 Pt 2):414-44.
28. Hausenloy DJ, Erik Botker H, Condorelli G, Ferdinandy P, Garcia-Dorado D, Heusch G, et al. Translating cardioprotection for patient benefit: position paper from the Working Group of Cellular Biology of the Heart of the European Society of Cardiology. Cardiovasc Res. 2013;98(1):7-27.
29. Vocci F, de With H. Consensus statement on evaluation of outcome of pharmacotherapy for substance abuse/ dependence: report from a NIDA/CPDD meeting. Bethesda, MD: National Institute on Drug Abuse Medications Development Division; 1999.
30. Labs KH, Dormandy JA, Jaeger KA, Stuerzebecher C, Hiatt WR. Trans-atlantic conference on clinical trial guidelines in PAOD (Peripheral arterial occlusive disease) clinical trial methodology. European Journal of Vascular & Endovascular Surgery. 1999;18(3):253-65.
31. Merkies IS, Lauria G. 131st ENMC international workshop: selection of outcome measures for peripheral neuropathy clinical trials 10-12 December 2004, Naarden, The Netherlands. Neuromuscul Disord. 2006;16(2):149-56.
32. Kenny KP, Day PF, Sharif MO, Parashos P, Lauridsen E, Feldens CA, et al. What are the important outcomes in traumatic dental injuries? An international approach to the development of a core outcome set. Dent Traumatol. 2018;34(1):4-11.
33. Feldman LS, Lee L, Fiore J, Jr. What outcomes are important in the assessment of Enhanced Recovery After Surgery (ERAS) pathways? Can J Anaesth. 2015;62(2):120-30.
34. Kloppenburg M, Boyesen P, Visser AW, Haugen IK, Boers M, Boonen A, et al. Report from the OMERACT Hand Osteoarthritis Working Group: Set of Core Domains and Preliminary Set of Instruments for Use in Clinical Trials and Observational Studies. J Rheumatol. 2015;42(11):2190-7.
35. Haywood KL, Griffin XL, Achten J, Costa ML. Developing a core outcome set for hip fracture trials. Bone & Joint Journal. 2014;96-B(8):1016-23.
36. Moniz-Cook E, Vernooij-Dassen M, Woods R, Verhey F, Chattat R, De Vugt M, et al. A European consensus on outcome measures for psychosocial intervention research in dementia care. Aging Ment Health. 2008;12(1):14-29.
37. Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int J Stroke. 2017;12(5):451-61.
38. Pushpanathan E, Setty T, Carvalho B, Sultan P. A Systematic Review of Postoperative Pain Outcome Measurements Utilised in Regional Anesthesia Randomized Controlled Trials. Anesthesiol Res Pract. 2018;2018:9050239.
39. Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain. 2018;159(3):481-95.

40. Spuls PI, Gerbens LAA, Simpson E, Apfelbacher CJ, Chalmers JR, Thomas KS, et al. Patient-Oriented Eczema Measure (POEM), a core instrument to measure symptoms in clinical trials: a Harmonising Outcome Measures for Eczema (HOME) statement. Br J Dermatol. 2017;176(4):979-84.
41. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH Task Force on research standards for chronic low back pain. J Pain. 2014;15(6):569-85.
42. Mindell JA, Emslie G, Blumer J, Genel M, Glaze D, Ivanenko A, et al. Pharmacologic management of insomnia in children and adolescents: consensus statement. Pediatrics. 2006;117(6):e1223-32.
43. Klokkerud M, Dagfinrud H, Uhlig T, Dager TN, Furunes KA, Klokkeide A, et al. Developing and testing a consensus-based core set of outcome measures for rehabilitation in musculoskeletal diseases. Scand J Rheumatol. 2018;47(3):225-34.
44. Allin BSR, Hall NJ, Ross AR, Marven SS, Kurinczuk JJ, Knight M, et al. Development of a gastroschisis core outcome set. Arch Dis Child Fetal Neonatal Ed. 2019;104(1):F76-F82.
45. Allin BSR, Bradnock T, Kenny S, Kurinczuk JJ, Walker G, Knight M, et al. NETS(1HD) study: development of a Hirschsprung's disease core outcome set. Arch Dis Child. 2017;102(12):1143-51.
46. Warners MJ, Hindryckx P, Levesque BG, Parker CE, Shackelton LM, Khanna R, et al. Systematic Review: Disease Activity Indices in Eosinophilic Esophagitis. Am J Gastroenterol. 2017;112(11):1658-69.
47. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129(4):424-47.
48. McNamara RL, Spatz ES, Kelley TA, Stowell CJ, Beltrame J, Heidenreich P, et al. Standardized Outcome Measurement for Patients With Coronary Artery Disease: Consensus From the International Consortium for Health Outcomes Measurement (ICHOM). J Am Heart Assoc. 2015;4(5).
49. Marrie RA, Miller A, Sormani MP, Thompson A, Waubant E, Trojano M, et al. Recommendations for observational studies of comorbidity in multiple sclerosis. Neurology. 2016;86(15):1446-53.
50. Carlson GA, Jensen PS, Findling RL, Meyer RE, Calabrese J, DelBello MP, et al. Methodological issues and controversies in clinical trials with child and adolescent patients with bipolar disorder: report of a consensus conference. J Child Adolesc Psychopharmacol. 2003;13(1):13-27.
51. Chow E, Wu JSY, Hoskin P, Coia LR, Bentzen SM, Blitzer PH. International consensus on palliative radiotherapy endpoints for future clinical trials in bone metastases. Radiother Oncol. 2002;64(3):275-80.
52. Bellm LA, Cunningham G, Durnell L, Eilers J, Epstein JB, Fleming T, et al. Defining clinically meaningful outcomes in the evaluation of new treatments for oral mucositis: oral mucositis patient provider advisory board. Cancer Investigation. 2002;20(5-6):793-800.
53. Meuleman C, Tomassetti C, D'Hooghe TM. Clinical outcome after laparoscopic radical excision of endometriosis and laparoscopic segmental bowel resection. Curr Opin Obstet Gynecol. 2012;24(4):245-52.
54. Schellinger PD, Bath PM, Lees KR, Bornstein NM, Uriel E, Eisert W, et al. Assessment of additional endpoints for trials in acute stroke - what, when, where, in who? Int J Stroke. 2012;7(3):227-30.
55. Sharrock AE, Barker T, Yuen HM, Rickard R, Tai N. Management and closure of the open abdomen after damage control laparotomy for trauma. A systematic review and meta-analysis. Injury. 2016;47(2):296-306.
56. van den Bos W, Muller BG, Ahmed H, Bangma CH, Barret E, Crouzet S, et al. Focal therapy in prostate cancer: international multidisciplinary consensus on trial design. Eur Urol. 2014;65(6):1078-83.
57. Wolters PL, Martin S, Merker VL, Gardner KL, Hingtgen CM, Tonsgard JH, et al. Patient-reported outcomes in neurofibromatosis and schwannomatosis clinical trials. Neurology. 2013;81(21 Suppl 1):S6-14.

58. Wallace SJ, Worrall L, Rose T, Le Dorze G, Breitenstein C, Hilari K, et al. A core outcome set for aphasia treatment research: The ROMA consensus statement. Int J Stroke. 2019;14(2):180-5.
59. Barlow GD, Lamping DL, Davey PG, Nathwani D. Evaluation of outcomes in community-acquired pneumonia: a guide for patients, physicians, and policy-makers. The Lancet Infectious Diseases. 2003;3(8):476-88.
60. Griffiths AM, Otley AR, Hyams J, Quiros AR, Grand RJ, Bousvaros A, et al. A review of activity indices and end points for clinical trials in children with Crohn's disease. Inflammatory Bowel Diseases. 2005;11(2):185-96.
61. van Riel PL. Provisional guidelines for measuring disease activity in clinical trials on rheumatoid arthritis. Br J Rheumatol. 1992;31(12):793-4.
62. Wilde EA, Whiteneck GG, Bogner J, Bushnik T, Cifu DX, Dikmen S, et al. Recommendations for the use of common outcome measures in traumatic brain injury research. Arch Phys Med Rehabil. 2010;91(11):1650-60 e17.
63. Grieve S, Perez R, Birklein F, Brunner F, Bruehl S, Harden RN, et al. Recommendations for a first Core Outcome Measurement set for complex regional PAin syndrome Clinical sTudies (COMPACT). Pain. 2017;158(6):1083-90.
64. Obbarius A, van Maasakkers L, Baer L, Clark DM, Crocker AG, de Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. Qual Life Res. 2017;26(12):3211-25.
65. Webster L, Groskreutz D, Grinbergs-Saull A, Howard R, O'Brien JT, Mountain G, et al. Core outcome measures for interventions to prevent or slow the progress of dementia for people living with mild to moderate dementia: Systematic review and consensus recommendations. PLoS One. 2017;12(6):e0179521.
66. Haywood K, Whitehead L, Nadkarni VM, Achana F, Beesems S, Bottiger BW, et al. COSCA (Core Outcome Set for Cardiac Arrest) in Adults: An Advisory Statement From the International Liaison Committee on Resuscitation. Resuscitation. 2018;127:147-63.
67. Hopkins C, Hettige R, Soni-Jaiswal A, Lakhani R, Carrie S, Cervin A, et al. 25 CHronic rhinosinusitis outcome MEasures (CHROME) – developing a core outcome set for trials of interventions in chronic rhinosinusitis. 25 Rhinology. 2018;56(1):22-32.
68. Fitzpatrick R, Chambers J, Burns T, Doll H, Fazel S, Jenkinso C, et al. A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. Health Technology Assessment 2010;14(18):1-94.
69. Khanna D, Lovell DJ, Giannini E, Clements PJ, Merkel PA, Seibold JR, et al. Development of a provisional core set of response measures for clinical trials of systemic sclerosis. Ann Rheum Dis. 2008;67(5):703-9.
70. Becker LB, Aufderheide TP, Geocadin RG, Callaway CW, Lazar RM, Donnino MW, et al. Primary outcomes for resuscitation science studies: a consensus statement from the American Heart Association. Circulation. 2011;124(19):2158-77.
71. Pavletic SZ, Martin P, Lee SJ, Mitchell S, Jacobsohn D, Cowen EW, et al. Measuring therapeutic response in chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: IV. Response Criteria Working Group report. Biol Blood Marrow Transplant. 2006;12(3):252-66.
72. Merkel PA, Herlyn K, Mahr AD, Neogi T, Seo P, Walsh M, et al. Progress towards a core set of outcome measures in small-vessel vasculitis. Report from OMERACT 9. J Rheumatol. 2009;36(10):2362-8.
73. Distler O, Behrens F, Pittrow D, Huscher D, Denton CP, Foeldvari I, et al. Defining appropriate outcome measures in pulmonary arterial hypertension related to systemic sclerosis: a Delphi consensus study with cluster analysis. Arthritis Rheum. 2008;59(6):867-75.
74. Miller FW, Rider LG, Chung YL, Cooper R, Danko K, Farewell V, et al. Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. Rheumatology 2001;40(11):1262-73.
75. Vargus-Adams JN, Martin LK. Measuring what matters in cerebral palsy: a breadth of important domains and outcome measures. Arch Phys Med Rehabil. 2009;90(12):2089-95.

76. Needham DM, Sepulveda KA, Dinglas VD, Chessare CM, Friedman LA, Bingham CO, 3rd, et al. Core Outcome Measures for Clinical Research in Acute Respiratory Failure Survivors. An International Modified Delphi Consensus Study. Am J Respir Crit Care Med. 2017;196(9):1122-30.
77. Gladman DD, Mease PJ, Strand V, Healy P, Helliwell PS, Fitzgerald O, et al. Consensus on a core set of domains for psoriatic arthritis. Journal of Rheumatology. 2007;34(5):1167-70.
78. Nikiphorou E, Mackie SL, Kirwan J, Boers M, Isaacs J, Morgan AW, et al. Achieving consensus on minimum data items (including core outcome domains) for a longitudinal observational cohort study in rheumatoid arthritis. Rheumatology (Oxford). 2017;56(4):550-5.
79. Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino MA, Bingham CO, III, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. J Rheumatol. 2014;41(5):1025-30.
80. Guyatt G, Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., Schunemann, H. J. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.
81. Vincent K, Kennedy S, Stratton P. Pain scoring in endometriosis: entry criteria and outcome measures for clinical trials. Report from the Art and Science of Endometriosis meeting. Fertil Steril. 2010;93(1):62-7.
82. Rief W, Burton C, Frostholm L, Henningsen P, Kleinstauber M, Kop WJ, et al. Core Outcome Domains for Clinical Trials on Somatic Symptom Disorder, Bodily Distress Disorder, and Functional Somatic Syndromes: European Network on Somatic Symptom Disorders Recommendations. Psychosom Med. 2017;79(9):1008-15.
83. Goldhahn J, Beaton D, Ladd A, Macdermid J, Hoang-Kim A, Distal Radius Working Group of the International Society for Fracture R, et al. Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. Arch Orthop Trauma Surg. 2014;134(2):197-205.
84. Wallace SJ, Worrall L, Rose T, Le Dorze G, Breitenstein C, Hilari K, et al. A core outcome set for aphasia treatment research: The ROMA consensus statement. Int J Stroke. 2018:1747493018806200.
85. Hellmich B, Flossmann O, Gross WL, Bacon P, Cohen-Tervaert JW, Guillevin L, et al. EULAR recommendations for conducting clinical studies and/or clinical trials in systemic vasculitis: focus on anti-neutrophil cytoplasm antibody-associated vasculitis. Ann Rheum Dis. 2007;66(5):605-17.
86. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49.
87. Ruemmele FM, Hyams JS, Otley A, Griffiths A, Kolho KL, Dias JA, et al. Outcome measures for clinical trials in paediatric IBD: an evidence-based, expert-driven practical statement paper of the paediatric ECCO committee. Gut. 2015;64(3):438-46.
88. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine. 2000;25(24):3100-3.
89. Micke O, Seegenschmiedt MH, German Working Group on Radiotherapy in Germany. Consensus guidelines for radiation therapy of benign diseases: a multicenter approach in Germany. International Journal of Radiation Oncology, Biology, Physics 2002;152(2):496-513.
90. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis and Rheumatism. 1993;36(6):729-40.
91. Ruiz CE, Hahn RT, Berrebi A, Borer JS, Cutlip DE, Fontana G, et al. Clinical Trial Principles and Endpoint Definitions for Paravalvular Leaks in Surgical Prosthesis: An Expert Statement. J Am Coll Cardiol. 2017;69(16):2067-87.
92. Stoner MC, Calligaro KD, Chaer RA, Dietzek AM, Farber A, Guzman RJ, et al. Reporting standards of the Society for Vascular Surgery for endovascular treatment of chronic lower extremity peripheral artery disease. J Vasc Surg. 2016;64(1):e1-e21.

93. Williams MR, Nayshtut M, Hoefnagel A, McKeown A, Carlson DW, Cravero J, et al. Efficacy Outcome Measures for Pediatric Procedural Sedation Clinical Trials: An ACTTION Systematic Review. Anesth Analg. 2018;126(3):956-67.
94. Hudak PL, Wright JG. The characteristics of patient satisfaction measures. Spine. 2000;25(24):3167-77.
95. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1-2):9-19.
96. Seegenschmiedt MH, Muller RP. Development of guidelines in radiation oncology. Strahlenther Onkol 1997;173:345-51.
97. Muller RP, Seegenschmiedt MH, eds. Leitlinien in der Radioonkologie. DEGRO: Eigenverlag Tubingen; 1999.
98. Hillner BE, Ingle JN, Berenson JR, al. e. American Society of Clinical Oncology guideline on the role of bisphosphonates in breast cancer. J Clin Oncol. 2000;18:1378-91.
99. Cook DL, Guyatt GH, Laupacis A, al. e. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest. 1989;102(Suppl 2):305S-11S.
100. Sackett DL. Rules of evidence and clinical recommendations on the use of anti-thrombotic agents. Chest. 1989;102(Suppl 2):2S-4S.
101. Bowman GP, Levine MN, Mohide EA, al. e. The practice guidelines development cycle: A conceptual tool for practice guidelines and implementation. J Clin Oncol. 1995;13:502-12.
102. Schmitt J, Apfelbacher C, Spuls PI, Thomas KS, Simpson EL, Furue M, et al. The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. J Invest Dermatol. 2015;135(1):24-30.
103. Tugwell P, Bombardier C. A methodologic framework for developing anfd selecting endpoints in clinical trials. J Rheumatol. 1982;9:758-62.
104. COSMIN Database of systematic reviews of outcome measurement instruments [cited 6 June 2019]. Available from: https://database.cosmin.nl/.
105. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57.
106. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.
107. COMET Initiative Database Search [cited 23 April 2019]. Available from: http://www.comet-initiative.org/studies/search.
108. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.
109. Ju A, Chow BY, Ralph AF, Howell M, Josephson MA, Ahn C, et al. Patient-reported outcome measures for life participation in kidney transplantation: A systematic review. Am J Transplant. 2019.
110. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. The Journal of rheumatology. 2019;46(8):1028-35.

Appendix A. Search strategy

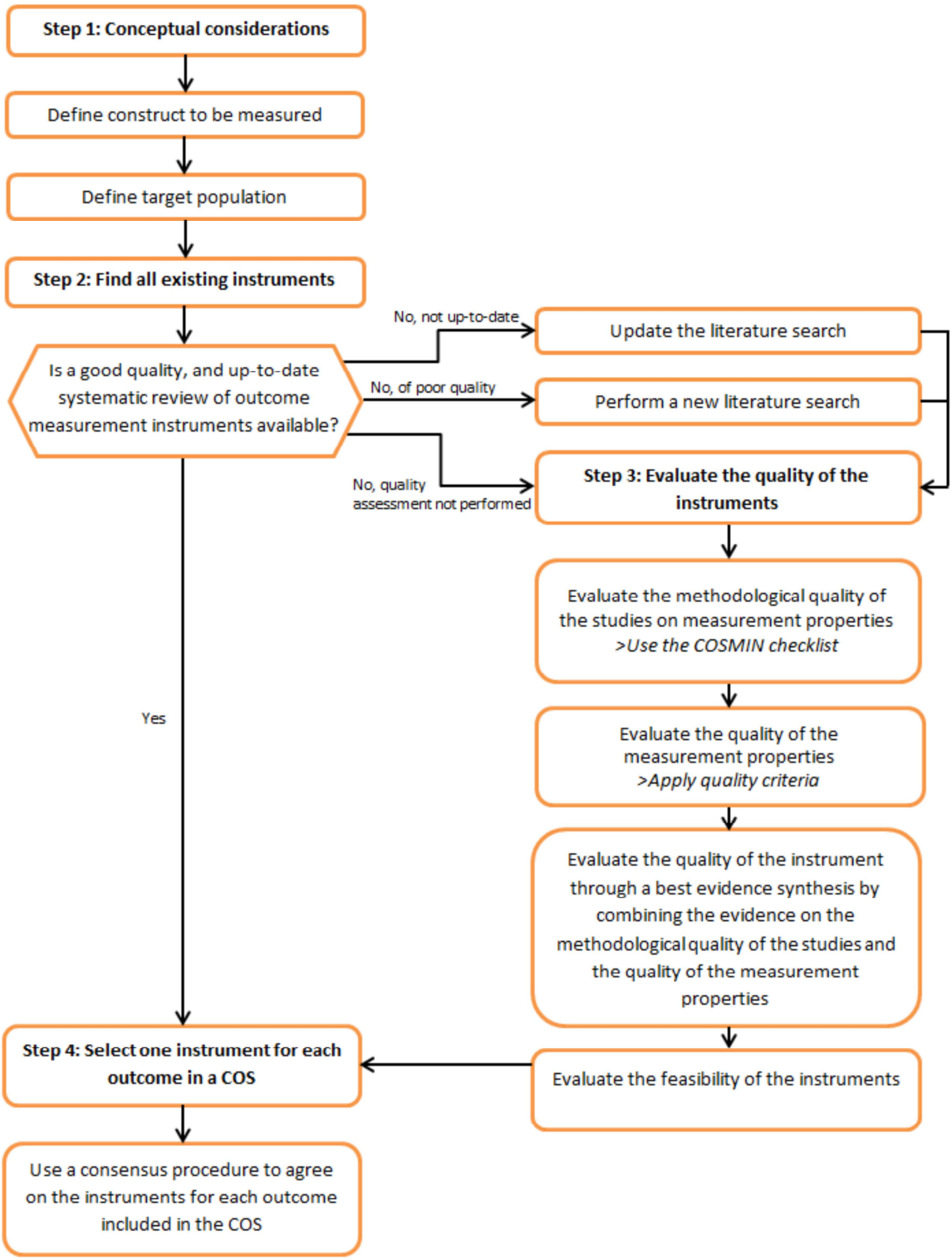
Search terms for MEDLINE	
	Randomised trial and systematic review terms
1	Health Services/ut [Utilization]
2	registries/
3	systematic review.mp.
4	structured review.ti.
5	evidence based medicine.ab.
6	exp Clinical Trials as Topic/
7	clinical trial$.ab.
8	randomised controlled trial$.ti,ab.
9	randomised trial$.ti,ab.
10	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9
	Methodology terms
11	workgroup$.mp.
12	standard$ outcome$.mp.
13	Practice Guideline/
14	clinical database.mp.
15	patient important outcome$.mp.
16	(standard$ adj3 reporting).mp.
17	congresses.pt.
18	Delphi Technique/
19	(recommend$ adj3 outcome$).mp.
20	consensus development conference.pt.
21	outcome$ reporting.mp.
22	priorit$ symptom$.mp.
23	(task force adj3 outcome$).mp.
24	appropriate outcome$.mp.
25	research design/
26	endpoint determination/
27	consensus development conference/
28	patient participation/
29	consensus.mp.
30	workshop.mp.
31	Consensus Development Conferences, NIH as Topic/
32	focus groups/

33	11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
	Outcome terms
34	outcome$.mp.
35	end point$.mp.
36	(core adj3 set).mp.
37	treatment emergent problem$.mp.
38	exp outcome Assessment Health Care/
39	Treatment Outcome/
40	Quality of Life/
41	34 or 35 or 36 or 37 or 38 or 39 or 40
	Key terms targeted
42	clinical-study design.mp.
43	patient$ perspective$.ti.
44	outcome$.mp. and delphi.ti.
45	(outcome$ and delphi).ab.
46	(perspective$ adj3 outcome$).ti.
47	core outcome$.ti,ab.
48	core set$.ti,ab.
49	clinical trial design$.ti.
50	design$ clinical trial$.ti.
51	(consensus and outcome$).ti.
52	42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51
53	10 and 33 and 41
54	52 or 53

Search terms for SCOPUS
((((INDEXTERMS(registries)) OR (INDEXTERMS(clinical trials as topic)) OR (ABS("evidence based medicine")) OR (ABS("clinical trial*")) OR (INDEXTERMS("Health Services Utilization")) OR (TITLE-ABS-KEY("SYSTEMATIC REVIEW")) OR (TITLE("structured review"))) OR (TITLE OR ABS("randomised controlled trial*")) OR (TITLE OR ABS (randomised trial*))) AND (((TITLE-ABS-KEY(workgroup*)) OR (TITLE-ABS-KEY(standard* outcome*)) OR (INDEXTERMS(practice guideline)) OR (TITLE-ABS-KEY("clinical database")) OR (TITLE-ABS-KEY("patient important outcome*")) OR (TITLE-ABS-KEY("standard* outcome*")) OR (INDEXTERMS(delphi technique))) OR ((TITLE-ABS-KEY(recommend* W/3 outcome*)) OR (TITLE-ABS-KEY(standard* W/3 reporting*)) OR (TITLE-ABS-KEY(task force W/3 outcome*)) OR (TITLE-ABS-KEY("appropriate outcome*")) OR (TITLE-ABS-KEY("outcome* reporting")) OR (TITLE-ABS-KEY("priorit* symptom*")) OR (INDEXTERMS(focus group)) (INDEXTERMS(research design))) OR ((INDEXTERMS(endpoint determination)) OR (INDEXTERMS(consensus development conference)) OR (INDEXTERMS(patient participation)) OR (TITLE-ABS-KEY(consensus)) OR (TITLE-ABS-KEY(workshop)))) AND 74) OR (((TITLE("design* clinical trials")) OR (TITLE(consensus AND outcome*)) OR

(TITLE-ABS-KEY("clinical-study design")) OR (TITLE("patient* perspective*")) OR (ABS(outcome* AND delphi)) OR (TITLE(outcome* AND delphi)) OR (TITLE(perspective* W/3 outcome*)) OR (ABS("core outcome*") OR TITLE("core outcome*"))) OR ((ABS("core set*") OR TITLE("core set*")) OR (TITLE("clinical trial design*"))))

Appendix B. Flowchart for the selection of outcome measurement instruments for core outcome sets



Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Terwee CB. Guideline for Selecting Outcome Measurement Instruments for Outcomes Included in a Core Outcome Set. 2016 [cited 7 April 2020]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-guideline-selecting-outcome-measurement-COS.pdf

Appendix C. Details of data extraction

Data extraction item	*Details*
Reference	
Reference no.	Consecutive number
Full reference	Full bibliographic information
Reference	First author; year
Extractor	Initials
Abstract	Full abstract
Study characteristics	
Geographical location (country) (locations of authors/ steering committee)	Free text
Geographical location (country) (locations of participants)	Free text
Language	Free text
Construct(s)	Free text
Disease area (COMET classification)	Free text
Target population	Free text
Age group	o Infants/children, adolescents o Adults o Not specified
Number of outcomes (what) for which an instrument (how) could be recommended	o Free text
Type of study	o What and how together o In two separate stages
Comments	o Free text
Methods for measurement selection	
Methods used to select core outcome measurement instruments	o finding existing instruments

	o quality assessment of instruments o recommendations on the selection of instruments o other, specify
Steps taken in finding existing instruments	o performing a systematic review, including search in MEDLINE (and EMBASE) o use existing review(s) o reference lists o expert opinion o other sources, specify
Quality assessment of the instruments	o evaluation quality of the study (i.e. COSMIN, EMPRO) o evaluation quality of the measurement properties (i.e. quality criteria applied) o other, specify o not done
Quality of evidence taken into consideration	o yes, specify o no o unclear, no information given
Feasibility aspects taken into consideration	o yes, specify o no o unclear, no information given
Arguments used for selecting instruments if no strict quality criteria were used	o Free text
Recommendations on the selection of instruments	o one instrument o multiple instruments for different subgroups o multiple instruments o other, specify o no instrument recommended o not reported
Consensus procedure used to agree on the instrument(s) included in the COS	o yes, specify o no

	o other, specify
COSMIN/COMET guideline used in selecting instruments?	o yes o no
Any guidance (i.e. published guidance) used in selecting instruments?	o yes, specify o no o unclear, no information given
Were the methods used for finding PROMs different vs. finding non-PROMs?	o yes, specify the methods used o no o not applicable
Recommendations for research on validation studies, new or additional instruments made?	o Yes, specify o no
Stakeholder groups involved in selecting instruments	o clinical experts o public representatives (including patients) o non-clinical research experts o authorities o industry representatives o others, specify o no details given
Comments on methods for outcome measurement selection	o Free text

Appendix D. Details on COS development studies (n=118)

Reference (1st author, year)	**Geographical location** (authors/ steering committee)	**Geographical location** (participants)	**Disease area** (Use classification of COMET)	**Target population**	**Age group** (1= infants/children, adolescents, 2= adults, 3= not specified)	**Number of outcomes** (what) for which an instrument (how) could be recommended	**Type of study** (1= what and how together, 2= in two separate stages)
Bombardier 2000	Canada	Canada	Orthopaedics & trauma	Spinal disorders	3	5	1
Khanna 2008	USA, Italy, UK	USA, Canada, South America, Asia, and Europe	Rheumatology	Systemic sclerosis (scleroderma)	3	11	2
Clements 2012	USA, France	USA, France	Rheumatology	Muskuloskeletal pain and systemic sclerosis (scleroderma)	2	9	1
Merkel 2009	USA, Germany, France, UK, the Netherlands, Turkey	USA, Germany, France, UK, the Netherlands, Turkey, and likely additional countries not listed	Rheumatology	Small-vessel vasculitis (ANCA-associated vasculitis)	3	3	2
Hellmich 2007	5 EU countries, USA (not further specified)	Germany, UK, the Netherlands, France, USA, Turkey	Rheumatology	Small-vessel vasculitis (ANCA-associated vasculitis)	1, 2	7	1
Douglas 2009	USA, Europe, Canada	USA, UK, Canada, Australia, Austria, Israel, Ireland, Spain, Denmark, Turkey, Korea, the Netherlands, Belgium, Iran	Endocrine & metabolic	Thyroid eye disease	3	14	1
Langguth 2007	Germany, New Zealand, Brazil, Sweden, USA, Italy, Belgium, Spain, France, the Netherlands, Colombia	Germany, New Zealand, Brazil, Sweden, USA, Italy, Belgium, Spain, France, the Netherlands, Colombia	Ear, nose & throat	Tinnitus	3	4	1
Clifton 1992	USA	USA, UK, the Netherlands	Neurology	Traumatic brain injury	2	8	1
Wood 1995	UK, USA, France, Canada, Belgium	UK, USA, France, Canada, Belgium	Infectious diseases	Herpes zoster	3	8	1
Hoeper 2004	Germany, USA, UK, Poland	not reported	Heart & circulation	Pulmonary arterial hypertension	3	6	1
Distler 2008	Switzerland, Germany, UK, France, Italy, Australia, USA	North America, Europe, Asia, Australia	Heart & circulation	Pulmonary arterial hypertension related Systemic Sclerosis	3	7	1
Penzien 2005	USA	USA, unknown	Neurology	Recurrent migraine and tension-type headache	1, 2	11	1
Becker 2011	USA	USA, France, Austria, Canada, Australia,	Heart & circulation	Cardiac arrest	3	5	1
Felson 1993	USA, the Netherlands, Canada	not reported	Rheumatology	Rheumatoid arthritis	3	7	2
Abellan van Kan 2011	France, USA	USA, France, uknown	Healthcare of older people	Sarcopenia	2 (older adults)	13 (prevention trials) 7 (intervention trials)	1
Gladman 2007	Canada, USA, UK, Australia, New Zealand, Italy, the Netherlands, Argentina, Ireland, Germany, Brazil	Canada, USA, UK, Australia, New Zealand, Italy, the Netherlands, Argentina, Ireland, Germany, Brazil, unknown	Rheumatology	Psoriatic arthritis	3	6	2
Dorman 2009	UK, USA, Australia	UK, USA, Australia	Health care of older people	Dyspnea or breathlessness in palliative care	3	8	1
Levine 2003	USA	USA	Urology	Peyronie's disease	3	5	1
Barlow 2003	UK	UK	Infectious diseases	Community-acquired diseases	3	5	1

Higashida 2003	USA	USA	Heart & circulation	Acute ischemic stroke	2	15	1
Miller 2001	US, UK, Hungary, Sweden, Czech Republic	USA, Hungary, Czech Republic, Sweden, UK, Israel, Mexico, France, Germany, the Netherlands, Canada, Guatemala, Italy, Korea	Rheumatology	Idiopathic inflammatory myopathies (IIM)	1,2	7	1
Canonica 2007	Italy, Argentina, France, US, Denmark, South Africa, Finland	Italy, Argentina, France, US, Denmark, South Africa, Finland	Lungs & airways	Respiratory allergy	3	7	1
Vellas 2008	France, Portugal, UK	France, Portugal, UK, Belgium, Denmark, the Netherlands, Switzerland, Italy, Germany, Sweden, USA	Neurology	Alzheimer's disease	3	2 (mild/early) 3 (moderate) 8 (severe)	1
Miller 1999	USA, UK	USA, UK, Canada, France, Sweden, the Netherlands, Swtizerland, Japan, Italy; 1998 conference uknown	Neurology	Amyotrophic lateral sclerosis/motor neurone disease	2	5	1
Leigh 2004	UK, Japan, Germany, France, USA, the Netherlands	UK, Japan, Germany, France, USA, the Netherlands, Switzerland	Neurology	Amyotrophic lateral sclerosis/motor neurone disease	3	3	1
Smith 1996	Australia	Australia, UK, USA	Lungs & airways	Asthma	1, 2	4	1
Reddel 2009	Australia, South Africa, USA, France, the Netherlands, UK, Canada, New Zealand	Australia, South Africa, USA, France, the Netherlands, UK, Canada, New zealand	Lungs & airways	Asthma	1, 2	7	1
Busse 2012	USA	USA	Lungs & airways	Asthma	1, 2	6 (age 5-11) 6 (age 12+)	1
Micke 2002	Germany	Germany	Other	Benign/ non-malignant disease	3	5	1
Carlson 2003	USA, Canada	USA	Mental health	Bipolar disorder	1	8	1
Chow 2002	Canada, UK, USA	Canada, USA, UK, the Netherlands	Cancer	Bone metastases	3	8	1
Partsch 2010	Austria, USA, Spain, Belgium, UK, Poland, Serbia, Hungary, Sweden, Germany, France, Australia	Austria, USA, Spain, Belgium, UK, Poland, Serbia, Hungary, Sweden, Germany, France, Australia, Italy, the Netherlands, Norway, Switzerland	Cancer	Breast cancer related lymphedema (BCRL)	3	18	1
Pallis 2011	Greece, UK, Belgium, Germany, the Netherlands, France, Italy, Switzerland	Greece, Belgium, UK, Germany, Italy, France, the Netherlands, Switzerland, Cyprus, Austria	Cancer	Solid tumors	2	6	1
Pavletic 2006	USA, Austria, Canada	USA, Austria, Canada	Blood disorders	Chronic graft-versus-host disease (GVHD)	1, 2	3	1
Vargus-Adams, 2009	USA	USA	Neurology; Child health	Cerebral palsy	1	8	2
Van Brussel, 2011	the Netherlands	the Netherlands	Child health	Children (and adolescents) with juvenile idiopathic arthritis, osteogenesis imperfecta, achondroplasia, hemophilia, cerebral palsy, spina bifida, cystic fibrosis, and childhood cancer	1	5	1
McGrath, 2008	Canada, USA, UK, Sweden	Canada, USA, UK, Sweden	Child health	Acute pain or chronic/recurrent pain	1	6 (acute) 8 (chronic/recurrent)	1

Griffiths, 2005	Canada, USA, UK	not reported	Gastroenterology	Crohn's disease	1	4	1
Ramsey, 1994	USA	USA	Lungs & airways	Cystic fibrosis	1, 2	8 (age <6 years) 10 (age >6 years)	1
Moniz-Cook, 2008	UK, the Netherlands, Italy, Ireland	France, Italy, the Netherlands, Ireland, Spain, Germany, Belgium, UK, Switzerland, Greece, Finland, Sweden, Poland	Neurology	Dementia	3	9	1
Capri, 1994	Italy, France, UK, Germany, USA, Spain	Italy, France, UK, Germany, USA, Spain	Lungs & airways	Chronic bronchitis and chronic obstructive pulmonary disease	3	3 (acute episodes) 3 (long-term treatment) 8 (symptomatic relief) 16 (modification of natural history of disease)	1
Meuleman, 2012	Belgium	Belgium	Gynaecology	(deep infiltrative) endometriosis	3	6	1
Schellinger, 2012	Germany, UK, USA, France	not reported	Heart & circulation	Acute stroke	3	5	1
Stout, 2012	USA, Austria, Hungary, Spain, Italy, UK, Belgium, the Netherlands, Australia, France, Germany	not reported	Skin; Heart & circulation	Lower extremity chronic edema receiving compression therapy.	2 (assumed)	14	1
Reneman, 2013	the Netherlands, Switzerland, USA	not reported	Orthopaedics & trauma	Chronic and subacute musculoskeletal pain	3	18	1
Donovan, 2012	USA	USA	Tobacco, drugs, & alcohol dependence	Drug dependence treatment	3	2	1
Vincent, 2010	UK, USA	UK, USA, Italy	Gynaecology	Endometriosis-related pain	3	8	1
Lamb, 2005	UK, Germany	Australia, Europe, North America	Orthopaedics & trauma	Community dwelling populations (prevention of injury associated with falls)	2	5	2
Fitzpatrick, 2010 (108 pages document)	UK	not reported	Mental health	Forensic mental health	1, 2	21	2
Alioum, 2001	France, the Netherlands, USA, UK	not reported	Gynaecology; Pregnancy & childbirth; Infectious disease	Postnatal transmission of HIV through breast milk	1	5	1
Tonetti, 2012	Italy, UK	not reported	Dentistry & oral health	Implant dentistry	3	8	2
Lux, 2004	UK	UK, Italy, USA, Argentina, Canada, Japan, Oman, Singapore, China, Philippines, Germany, Cuba, Switzerland, Malaysia, Thailand	Child health; Neurology	Infantile spasms and West Syndrome (Epilepsy)	1	9	2
Mindell, 2006	USA	USA	Neurology	Insomnia	1, 2	4	2
Giannini, 1997	USA, Italy	Sweden, Hungary, USA, Spain, Canada, Italy, the Netherlands, Portugal, Germany, France, Finland, Russia, Belgium, UK	Rheumatology	Juvenile arthritis	1	6	2
Ruperto, 2003	Italy, Australia, USA, Sweden, Canada, Argentina, the Netherlands, Brazil, France, Greece, UK	Italy, Australia, USA, Sweden, Canada, Argentina, the Netherlands, Brazil, France, Greece, UK, South Korea, Hungary, Mexico, Spain, Czech Republic, Croatia, Israel, Switzerland, Finland, Austria, Germany,	Rheumatology	Juvenile systemic lupus erythematosus and juvenile dermatomyositis	1	7 (JSLE: 4 disease activity; 3 disease damage) 11 (JDM: 6 disease activity; 5	2

		Belgium, Norway, Portugal, Bulgaria, Denmark, Russia, Turkey, Georgia, Poland, Latvia, Yugoslavia, Slovakia				disease damage)	
Deyo, 1998	USA, Canada, the Netherlands, UK, Finland	USA, Canada, the Netherlands, UK, Finland	Orthopaedics & trauma	Low back pain	3	5	1
Porst, 2010	Germany, Israel, Turkey, USA, Korea, Brazil, UK	Germany, Israel, Turkey, USA, Korea, Brazil, UK	Urology	Male sexual dysfunctions (erectile disfunction, premature ejaculation, delayed/absent ejaculation, libido disorders/loss of desire, hypogonadism and Peyronie's disease)	3	2 (clinical trials in ED) 5 (ED PROs) 3 (clinical trials with a gene transfer in ED) 2 (clinical trials in PE) 2 (clinical trials in hypogonadism)	1
Sanyal, 2011	USA, France	not reported	Endocrine & metabolic	Non-alcoholic steatohepatitis	1, 2	9 (at risk of progression to cirrhosis) 4 (cirrhosis) 2 (NASH following liver transplantation) 2 (children) 1 (NASH in children) 2 (safety-related endpoints in NASH)	1
Bellm, 2002	USA, Canada	not reported	Dentistry & oral health; Cancer	Oral mucositis	3	6	2
Heiligenhaus, 2012	Germany, UK, USA, Switzerland, France, the Netherlands, Spain, Finland, Denmark	not reported	Rheumatology	Juvenile idiopathic arthritis (JIA)–associated uveitis	1	11	2
Hausenloy 2013	UK, Denmark, Italy, Hungary, Spain, Germany, South Africa, the Netherlands, USA, France	N/A	Heart & circulation	At risk of acute myocardial ischaemia-reperfusion injury	3	7 (STEMI patients) 6 (CABG patients)	1
Keim 2004	USA, Canada	not reported	Lungs & airways	Respiratory distress in the out-of-hospital setting	1, 2	2	2
Tfelt-Hansen 2012	Denmark, Spain, USA, Sweden, Italy, Germany, France	Denmark, Spain, USA, Sweden, Italy, Germany, France	Neurology	Acute migraine attacks and migraine porphylaxis	1, 2	11 (acture migraine)	1
Lipton 1995	USA	not reported	Neurology	Cluster headache	2	6 (migraine prophylaxis)	1
Bendtsen 2009	Denmark (and likely others, but no details for affiliations of co-authors given)	not reported	Neurology	Tension-type headache	2	1 (children)	1
Vocci 1999	USA	USA (possible others)	Tobacco, drugs & alcohol dependence	Smoking cessation, alcohol abuse and dependence, cocaine abuse	1 , 2 (assumed)	5 (acute treatment)	1
van Riel 1992	the Netherlands	N/A	Rheumatology	Rheumatoid arthritis	3	7	2
Labs 1999	USA, UK, Switzerland, Germany	Switzerland, UK, Sweden, USA, Italy, Germany, Canada, Belgium, Greece, Ireland, Austria, France	Heart & circulation	Peripheral arterial occlusive disease	3	6 (intermittent claudication) 3 (critical limb ischemia)	1

Merkies 2006	the Netherlands	France, Italy, the Netherlands, UK, USA	Neurology	Peripheral neuropathy	3	4 (painful neuropathy) 3 (GBS, CIDP, MGUSP, MMN, CMT) 3 (diabetic neuropathy)	1
Reilly 2006	UK, Belgium, Italy	Austria, Belgium, Canada, Czech Republic, France, Germany, Italy, the Netherlands, Spain, UK, USA	Neurology	Charcot-marie-tooth disease type 1 A	2	4	1
Leon 2011	the Netherlands, USA, France, Canada, Switzerland	the Netherlands, USA, Germany, Canada, Switzerland, France, UK	Heart & circulation	Transcatheter aortic valve implantation	3	16	1
Wilde 2010	USA, Germany	US, Germany	Neurology	Traumatic brain injury	2	8	1
Chen 2014	USA, Canada	not reported	Cancer	Localized and advanced prostate cancer	2	5 (localized prostate cancer) 4 (advanced cancer)	1
Diehm 2013	Switzerland, Belgium, the Netherlands	Switzerland, Belgium, the Netherlands, USA (potentially other European countries as well)	Heart & circulation	Aortic dissection	2	21	1
Feldman 2015	Canada	Canada	Rehabilitation	Enhanced recovery after surgery pathways following major abdominal surgery	3	5 (intermediate recovery phase: in hospital) 4 (late recovery phase: post-discharge)	1
Goldhahn 2014	Switzerland, Canada, USA	Switzerland, Canada, UK, Belgium, USA, the Netherlands, Italy, France, Ireland	Orthopaedics & trauma	Distal radius fractures	2	5	1
Haywood 2014	UK	UK	Orthopaedics & trauma	Hip fracture	2	5	1
Kloppenburg 2015	the Netherlands, Norway, UK, Canada	not reported	Rheumatology	Hand osteoarthritis	3	6 (all settings) 8 (trials of symptom modification and observational studies)	2
Wolters 2013	USA	USA, UK	Genetic disorders	Neurofibromatoses	1, 2	4	1
Ruemmele 2014	France, USA, Canada, Finland, Portugal, Israel, the Netherlands, Hungary, Belgium, UK	France, USA, Canada, Finland, Portugal, Israel, the Netherlands, Hungary, Belgium, UK	Gastroneterology	Crohn's disease and ulcerative colitis	1	6	1
Khanna 2015	USA, Canada, Australia, New Zealand, UK	USA, Canada, Australia, New Zealand, UK, South Africa, France, Greece, Portugal, Germany, Italy, the Netherlands, Spain, Brazil, Argentina, Uruguay, Hungary, Belgium, Switzerland, Sweden, Bulgaria, Norway, Finland, Denmark, Korea, Turkey, Poland, Japan, India, Pakistan, Ukraine, Mexico, Chile, Ireland	Lungs & airways	Connective tissue diseases-associated interstitial lung disease	3	7	1
Ball 2013	UK	UK	Orthopaedics & trauma	Dupuytren's disease	2	5	1

Paul 2014	UK, Ireland, New Zealand, USA	UK, Ireland, New Zealand, USA	Neurology	Multiple Sclerosis	3	7	1
van den Bos 2014	the Netherlands, UK, France, USA, Belgium, Germany	UK, Spain, USA, the Netherlands, France, Italy, Ireland, Austria, Germany, Belgium, Canada	Cancer	Localised prostate cancer	2 (assumed)	4	1
Van den Bos 2015	the Netherlands, USA, Germany, Canada, UK, Switzerland, France	the Netherlands, USA, Germany, Canada, UK, Switzerland, France, Norway, Italy, Austria, Japan, Belgium	Cancer	Localised prostate cancer	3	8	1
Deyo 2014	USA	USA	Orthopaedics & trauma	Chronic lower back pain	2 (assumed)	6	1
Pinder 2015	UK	UK	Orthopaedics & trauma	Scaphoid nonunion fracture	2	4	2
Allin 2017	UK	UK and International participants (countries not reported)	Gastroenterology; Child health	Hirschsprung's disease in high-income countries	1	10	2
Grieve 2017	UK, the Netherlands, Germany, Switzerland, USA, Canada, Denmark, Israel	Switzerland, Germany, the Netherlands, Canada, USA, Denmark, Norway, Israel, Australia, New Zealand, South Africa, Japan, Argentina, Brazil, UK	Anaesthesia & pain control	Chronic regional pain syndrome	2	7	2
Hernandez Yenty 2016	the Netherlands	the Netherlands	Other	Congenital or acquired inverted nipples in females	2	6	1
Kenny 2018	UK, Australia, Denmark, Brazil, USA, Canada, Sweden, China, Japan, Kuwait	UK, Australia, Denmark, Brazil, USA, Canada, Sweden, China, Japan, Kuwait; E-Delphi completed by dentists in over 30 countris but does not provide information	Dentistry & oral health	Traumatic fental injury	1, 2	13 (all injury types) 10 (injury specific)	2
Klokkerud 2017	Norway	Norway	Musculoskeletal Disease	MSD undergoing rehabilitation	2 (assumed)	10	2
Marrie 2016	Canada, Italy, UK, USA	USA, UK, Germany, Canada, France, Denmark, Sweden, the Netherlands	Neurology	Relapsing remitting MS or primary progressive MS	3	6	1
McNamara 2015	USA, Australia, Spain, Sweden, Singapore, India, UK	USA, Australia, Sweden, Singapore, India, UK	Heart & circulation	Coronary artery disease	3	13	1
Nabbout 2018	France, UK	France, Italy, USA, Australia, UK	Neurology; Child health	Dravet syndrome	1, 2	5	1
Nikiphorou 2017	UK, the Netherlands	UK	Rheumatology	Rheumatoid arthritis	2 (assumed)	10	1
Obbarius 2017	Germany, USA, UK, Canada, the Netherlands, Saudi Arabia, Japan, Sweden, Australia, Uganda, Brazil, Chile	Australia, Brazil, Canada, Chile, Germany, India, Japan, the Netherlands, Sweden, Uganda, UK, USA	Mental health	Depression and/or anxiety	1, 2	10	1
Ong 2017	Australia, USA, the Netherlands, Ireland, Belgium, Mexico, Sweden, Malaysia	Australia, USA, the Netherlands, Ireland, Belgium, Mexico, Sweden, Malaysia	Cancer	pathologically confirmed American Joint Committee of Cancer (AJCC) patients with stages 0 to IVBC	2	26	1
Rief 2017	Germany, UK, Denmark, the Netherlands, Norway	not reported	Mental health	Somatic symptoms and associated disorders	3	11	1
Ruiz 2017	USA, Italy, Canada, Spain, UK,	not reported	Heart & circulation	PVL closure devices	3	8	1
Sharrock 2016	UK	N/A	Orthopaedics & trauma	Abdominal closure following damage control laparotomy for trauma	1, 2 (assumed)	5	2
Stoner 2016	USA	Italy, UK	Heart & circulation	Peripheral arterial disease	2 (assumed)	5	1
Needham 2017	USA	USA, UK, Australia, Canada, Singapore, China, France, Germany, Belgium, Greece,	Rehabilitation; Lungs & airways	Acute respiratory failure survivors	2 (assumed)	8	2

		the Netherlands, Norway, Italy, Ireland, Crazil, Panama					
Wallace 2019	Australia, Canada, Germany, UK, USA, Ireland, Sweden	Australia, Canada, Germany, UK, USA, Ireland, Sweden	Neurology	Post-stroke aphasia	2	5	2
Warners 2017	the Netherlands, Canada, Belgium, USA	N/A	Gastroenterology	Eosinophilic esophagitis disease	1, 2	4	2
Webster 2017	UK, Germany	UK	Neurology	Mild or moderate dementia	2	2	1
Williams 2018	USA, France, Israel, China	USA, France, Israel, China	Anaesthesia & pain control	Pediatric sedation procedures	1	4	2
Dohner 2017	Germany, USA, UK, Italy, France, Japan, the Netherlands, Spain, Taiwan, Australia	Germany, USA, UK, Italy, France, Japan, the Netherlands, Spain, Taiwan, Australia	Cancer	Acute myeloid leukemia	2	4	1
Kwakkel 2017	USA, Australia, UK	not reported	Neurology; Rehabilitation	Sensorimotor recovery after stroke	2	7	1
Allin 2018	UK	UK + other high income countries (not reported)	Gastroenterology	Gastroschisis	1	8	2
Balakrishnan 2018	USA, Argentina, Canada, France, Switzerland, Ireland, Italy, Chile, Australia	8 countries within North and South America, Europe, and Australia (not reported)	Lungs & airways	Pediatric airway reconstruction	1	8 (General) 13 (Site-Specific for Laryngeal Reconstruction - Supraglottis) 15 (Site-Specific for Laryngeal Reconstruction - Glottis) 17 (Site-Specific for Subglottic Laryngeal Reconstruction) 8 (Site-Specific for Cervical Tracheal Reconstruction) 12 (Site-Specific for Thoracic Tracheal Reconstruction)	1
Haywood 2018	USA, UK, the Netherlands, Germany, Finland, Sweden, Belgium, Australia, Canada, Singapore	UK, the Netherlands, Finland, Germany, Belgium, Sweden, USA, Canada, Singapore, Australia, New Zealand + 4 others	Heart & circulation	Cardiac arrest	2	3	2
Hopkins 2018	UK, Australia, New Zealand, the Netherlands, Belgium, Germany, USA, Singapore, Canada	UK, USA	Ear, nose & throat	Chronic rhinosinusitis	2	15	1
Iorio 2018	Canada, USA	North America, Europe, possbily others (countries not reported)	Blood disorders	Haemophilia	1, 2 (assumed)	6	1
Pushpanathan 2018	UK, USA	N/A	Anaesthesia & pain control	Postoperative pain	2	6	1
Radner 2018	Austria, Sweden, UK, France, Romania, the Netherlands, Germany, Spain, Switzerland, Norway, Denmark, USA, Portugal,	28 different European countries (not reported)	Rheumatology	Rheumatoid arthritis	2 (assumed)	6	1

	Czech Republic						
Chiarotto 2018	the Netherlands, USA, Australia, Brazil, UK, Norway, Spain	USA, the Netherlands, Australia, UK, Brazil, Italy, Norway, Canada, Spain, Germany, Finland, Denmark, Switzerland + others	Orthopaedics & trauma	non-specific low back pain	2	4	2
Spuls 2017	Germany the Netherlands, UK, Brazil, Israel, USA, Japan, France, Australia, Sweden	Germany, the Netherlands, UK, Brazil, Israel, USA, Japan, France, Australia, Sweden, Canada, Denmark, China, Tanzania	Skin	Eczema	1, 2	4	2

What is new?

Key findings

- Methods used to select core outcome measurement instruments vary across studies, with many studies not meeting the recommended standards.
- Methods used to select outcome measurement instruments have improved since the publication of the COSMIN/COMET guideline.

What this adds to what is known?

- This is the first study to assess how the outcome measurement instruments recommended in existing core outcome sets have been selected and whether good practices are being followed.

What is the implication, what should change now?

- Core outcome set developers need to make better use of the guidance available when agreeing on how to measure the outcomes included in core outcome sets.
- Developers need to ensure that outcome measurement instruments are of sufficient quality, especially have sufficient content validity.

Author statement:

Sarah L Gorst: Investigation; Methodology; Roles/Writing – original draft; Writing – review & editing; Project administration.

Cecilia AC Prinsen: Investigation; Methodology; Roles/Writing – original draft; Writing – review & editing; Project administration.

Maximilian Salcher-Konrad: Investigation; Writing – review & editing.

Karen Matvienko-Sikar: Investigation; Writing – review & editing.

Paula R Williamson: Conceptualization; Methodology; Writing – review & editing.

Caroline B Terwee: Conceptualization; Methodology; Writing – review & editing.

Conflict of interest statement: PRW is a member of the COMET (Core Outcome Measures in Effectiveness Trials) Management Group. CBT is founder of the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) initiative. The remaining authors declare no conflict of interest.